

Fehérje-térszerkezetek homológia modellezése

MTA Doktori Értekezés

2008 szeptember

Fiser András

MTA Szegedi Biológiai Központ Enzimológiai Intézete

Budapest, Karolina út 29.

Department of Systems and Computational Biology

Department of Biochemistry

Albert Einstein College of Medicine, New York, USA

1. BEVEZETÉS

Szekvenciák, szerkezetek, szerkezeti genomika

A fehérjék funkcionális jellemzése az egyik leggyakoribb feladat a molekuláris biológiai kutatásban. A fehérje szekvenciája fontos információforrás, azonban a fehérjeszekvenciák azon tulajdonsága, hogy néha egymástól nagyon eltérő szekvenciák is képesek ugyanazt a funkciót ellátni, gyakran lehetetlenné teszi a funkcionálisan fontos aminosavak azonosítását. Például enzimeknél a 40% feletti szekvencia azonosság általában hasonló funkcióra utal, de a 30-40%-os szekvencia azonosság már csak az első három EC (Enzyme Commission) azonosító szám meghatározását tudja garantálni és csak 90% biztonsággal. Amíg 30% alatti szekvencia azonosság esetén már térszerkezeti információ szükséges, illetve nélkülözhetetlen a pontos funkció megállapításához. Mindeközben a homológ enzimek 75%-a kevesebb mint 30%-os szekvencia azonossággal rendelkezik.

A szerkezet alapú fehérjefunkció jellemzés széles skálát ölel fel. Olyan alapfokú információktól kezdve, mint például a helyes fold meghatározása, ami gyakran egy általános funkcióra jellemző, egészen olyan magas szintű jellemzésig, mint a szubsztrátspecifitás meghatározása vagy a szerkezet alapú inhibitor tervezés.

A genom szintű szekvenálás segítségével már több mint 6 millió egyedi fehérjeszekvenciát határoztak meg, és ez a szám megduplázódhat, ha a Craig Venter által irányított Globális Tengeri Felderítési program által összegyűjtött metagenomikai adatokat is publikálják /J.C. Venter és mtsi. Science 2004, 304, 66/. Ugyanakkor csak körülbelül 50 000 fehérje térszerkezete ismert, amelyeket röntgen krisztallográfiával vagy NMR-el oldottak meg. Mivel ezek a kísérleti térszerkezet meghatározási módszerek sokkal drágábbak és időigényesebbek, mint a fehérjeszekvenálás, a térszerkezeti is jellemzett fehérjék hányada a mai körülbelül mindössze 1%-ról még tovább fog csökkenni.

Az ezredforduló táján szerkezetgenomikai projektek indultak be világszerte. Ezen erőfeszítések célja, hogy néhány ezer megfelelően kiválasztott fehérjének a térszerkezetét megoldják. A megoldott szerkezetek ezután mintaként szolgálnak (templát) az egész fehérjecsalád modellezésére. Ezen kutatások azon a felismerésen alapulnak, hogy a ma ismert több millió szekvencia feltehetően néhány ezer jól megkülönböztethető fold-szerkezetbe tekeredik fel. E sejtés alapja, hogy az újonnan meghatározott szerkezetekről egyre gyakrabban derül ki, hogy egy már ismert fold újabb reprezentásai. A szerkezetgenomikai projektek mára a fehérje térszerkezetek meghatározó forrásává váltak. Az elmúlt években a Fehérje Adatbankba (PDB) elhelyezett egyedi szerkezetek 75%-a szerkezetgenomikai kísérletekből származott.

Ugyanakkor a szerkezetgenomikai megközelítés is tovább erősíti az igényt hatékony szerkezetmodellező módszerek kifejlesztésére, hiszen ezen projekteken belül a megoldott egyedi templátszerkezeteket használják fel a megcélzott fehérjecsaládok térszerkezetének teljes feltérképezésére.

Szerkezetmodellező módszerek

A fehérjék szerkezetét szabályozó törvényszerűségeket vagy a fizika törvényein vagy az evolúción keresztül lehet megközelíteni. E két alapvető megközelítési mód az alapja a két fő fehérjemodellező iskolának.

Az egyik megközelítés, az *ab initio* vagy templátmentes modellező módszer csak szekvenciális információra alapszik. Az *ab initio* módszerek feltételezik, hogy a fehérje natív szerkezete megfelel a globális energiaminimum állapotának és arra törekednek, hogy ezt a szerkezetet megtalálják számtalan konformáció feltérképezésével és energiájuk kiszámolásával.

A másik megközelítési mód, amit templátfüggő modellezésnek is hívnak, az ún. threading-et és a homológia- vagy összehasonlító fehérje-modellezést foglalja magába. Ez a típusú modellezés akkor kivitelezhető, ha a modellezni kívánt szekvencia nagy része detektálható hasonlóságot mutat egy már ismert szerkezetű fehérjeszekvenciával. Amint egy fehérjecsaládban legalább egy szerkezetet kísérletesen megoldanak, a többi tag szerkezete homológia modellezéssel előállítható a köztük meghatározott szekvencia-illesztés alapján. A homológia modellezés azért lehetséges, mert egy kis mértékű szekvenciaváltozás általában kis mértékű szerkezeti változást okoz, és mert a fehérjék szerkezete jobban konzerválódott az evolúció során, mint a szekvenciája. Emiatt, ha két fehérje között szekvenciaszintű hasonlóság állapítható meg, akkor az nagy valószínűséggel szerkezeti hasonlóságra is utal. A homológia modellezés egyre bővülő alkalmazása annak köszönhető, hogy a lehetséges fehérjeszerkezeti típusok, ún. foldok száma erősen korlátozott, és hogy kísérleti úton egyre nagyobb hányadának válik ismertté egy-egy képviselője.

Mindkét szerkezetmodellező megközelítésnek vannak az előnyei és hátrányai. Elvileg az *ab initio* módszer bármilyen szekvenciára alkalmazható, azonban a gyakorlatban a fehérje folding probléma bonyolultsága és korlátozott ismeretünk a folding folyamatáról azt eredményezi, hogy viszonylag rossz felbontású, pontatlan modelleket lehet csak építeni és az alkalmazhatóság kisméretű, 100 aminosavnál rövidebb fehérjékre korlátozódik. A fehérjemodellezők két évente megrendezésre kerülő CASP konferenciája az egyik objektív

fórum a módszerek tesztelésére, mert kísérletesen is megoldott, de szándékosan nem publikált fehérjék térszerkezetét kell modellezni a résztvevőknek. A konferencia szakmai szervezői különböző módszerekkel összevetik a benyújtott modelleket a kísérletesen meghatározott szerkezetekkel és rangsorolják őket a pontosságuk alapján, és így az elkészítésükhöz használt módszereket is. A 2007-es CASP konferencia szerint *ab initio* módszerek az esetek nagyobb részében még mindig nem képesek a pontos fold meghatározására. Az *ab initio* módszerekkel szemben a homológia modellezés általában olyan minőségű modelleket képes építeni, amik összevethető pontosságúak az alacsony felbontású röntgenkristallográfiás vagy közepes felbontású NMR szerkezetekkel. Jelenleg kb. 70% annak a valószínűsége, hogy egy véletlenül kiválasztott szekvencia legalább egy doménjához találunk egy hasonló kísérleti szerkezetet.

A gyakorlatban a templatfüggő modellező eljárások (pl. Modeller program) mindig használnak olyan információt, ami független a templatától, pl. különböző erőterek formájában, amiket statisztikai vizsgálatokból vagy molekulamechanikai számításokból lehet megkapni. A folyamatosan javuló erőtereknek és mintavételező algoritmusoknak köszönhetően a legsikeresebb homológia szerkezetmodellező eljárások egyre gyakrabban térképeznek fel a templat szerkezettől független konformációkat. Hasonló módon, a legsikeresebb *ab initio* eljárások (pl. Rosetta program) valójában ismert fehérjék rövid fragmenseit használják fel a szimulációk során. Általában, a klasszikus *ab initio* eljárások a fehérje folding folyamatába engednek bepillantást nyerni, míg az effektív molekulamodellező módszerek mindig valamilyen templatfüggő eljárások.

Fehérjeszerkezetek homológia modellezése

E módszer szükséges feltétele, hogy (i) a modellezett szekvencia hasonlóságot mutasson legalább egy templat szerkezethez a szekvencia döntő részében, és (ii) hogy egy pontos szekvencia-illesztést lehessen kapni köztük. A homológia alapú fehérjemodellezés négy általános lépésben összegezhető: (1) a megfelelő templat szerkezetek azonosítása és kiválasztása (2) a szekvencia és templat szerkezet illesztésének kiszámítása (3) a modell megépítése és (4) a molekulamodell minőségének kiértékelése.

A homológia modellezéssel nyert fehérjemodellek leggyakoribb hibáit öt kategóriába lehet sorolni. (1) Helytelen oldallánc konformációk; (2) szerkezeti eltérések vagy torzulások az amúgy a szekvencia szempontjából helyesen illesztett régiókban (ún. merev test torzulások); (3) szerkezeti eltérések vagy torzulások olyan régiókban, amelyeknek nincs illeszthető szegmense a templat szerkezetben (loop vagy hurok modellezés); (4) szerkezeti elmozdulások vagy torzulások helytelenül szekvencia-illesztett régiókban (szekvencia-illesztési hiba); (5)

helytelenül feltekeredő szerkezet helytelen templátazonosítás miatt. A 3-5 hiba típusok viszonylag ritkák, ha körülbelül 40% vagy a feletti szekvenciaazonosság áll fenn a modellezett szekvencia és a templát között. Ilyen esetekben, minőségi szempontból, a peptidváz atomjainak a ~90%-a 1Å RMSD-n (Root Mean Square Deviation, az eltérések négyzetösszegének a gyöke) belül lesz a végső modellben, a hibák az inzerciók (loop szegmensek) és az oldalláncok modellezésére korlátozódnak. Ugyanakkor, ha a szekvencia azonosság 30% alá csökken, akkor a szekvencia-illesztési probléma válik döntővé. Ez átlagosan az aminosavak 20%-a helytelen illesztését, és következésképpen rossz modelljét eredményezi, így a végső szerkezeti modell csak kb. 3Å RMSD pontosságú lesz. A szekvencia-illesztési hibák ebben a tartományban komoly akadályt jelentenek a modellező módszerek alkalmazásában, mert a természetben ismert, amúgy szerkezetileg homológ fehérjék döntő része kevesebb mint 30%-os szekvencia azonosságot mutat egymással.

A homológia modellezés elvét már a '80-as évek vége felé felvetették és az első széles körben elterjedt módszerek a 90-es évek elején-közepén érték a szélesebb tudományos közeget. Ugyanakkor az ekkor elérhető módszereket csak meglehetősen egyszerű esetekre lehetett alkalmazni, amikor a modellezett szekvencia és a templát triviálisan hasonlóak voltak, kb. több mint 40%-os szekvencia azonosságot mutatva. Ez viszont a modellezés alkalmazhatóságát a természetben előforduló esetek töredékére korlátozta. A nagyméretű szekvenálási programokkal párhuzamosan nagy igény alakult ki hatékony és széles körben alkalmazható fehérjemodellező eljárások kifejlesztésére. Munkám során a homológia modellezés minden aspektusát érintő új eljárásokat dolgoztam ki.

2. CÉLKITÜZÉSEK

Az elmúlt 11 évben az elméleti biofizika területén belül azzal foglalkoztam, hogy új algoritmusokat dolgozzak ki fehérjemolekulák térszerkezetének modellezésére, és hogy kísérletes csoportokkal együttműködve alkalmazzam ezen módszereket specifikus biológiai problémák vizsgálatára.

2.1. Szerkezet alapú szekvencia-illesztési módszer

A meglévő szekvencia-illesztési módszerek egymásnak ellentmondó eredményt produkálnak, ha az illesztendő fehérjeszekvenciák kevesebb mint kb. 20-30%-ban azonosak. Ugyanakkor a homológia modellezés legkritikusabb lépése a pontos szekvencia-

illesztés meghatározása. Ezért új, hatékony szekvencia-illesztő módszert dolgoztunk ki távoli rokonságot mutató fehérjékre, azt kihasználva, hogy molekulamodellezés esetén az illesztett fehérjék egyikének a térszerkezete mindig ismert kell, hogy legyen.

2.2. Loopok modellezése

A homológia modellezés gyakran nem tud szerkezeti modellt nyújtani szekvencia inzerciókra, amelyek a térszerkezetben leggyakrabban a molekula felszínén található loopoknak felelnek meg, mert ezek szekvenciája különbözik a modellezett fehérje és a templát között. Az, hogy az eltérések a leggyakrabban a loopokban fordulnak elő nem véletlen, hiszen ezek gyakran felelősek a fehérje egyedi funkciójáért. A loopok stratégiai fontossága miatt két új loop- modellező módszert is kifejlesztettünk, ahol az egyik *ab initio*, a másik pedig fragmens kereső eljárást használ.

2.3. Célzott mutációk modellezése

Az egyik leggyakoribb kísérletes szerkezet/funkció manipuláló eszköz a célzott mutációk alkalmazása. Amíg az 1-2 Å RMSD pontosságú modell megfelelő betekintést ad a teljes szerkezet szerveződésébe, addig egy specifikus oldallánc módosítása esetén kíváncsi, hogy pontosan modellezhető legyen az oldallánc, ami akár tized angstromos RMSD pontosságot is megkívánhat. Új, nagy pontosságú, célzott mutációmodellező eljárást fejlesztettünk ki, ami ötvözi az *ab initio* eljárásokat a templát szerkezetből kinyerhető információval.

2.4. Statisztikai potenciálok fejlesztése

Minden térszerkezet-modellező algoritmus, oldallánc vagy loop-modellezés valamint szerkezet alapú szekvencia-illesztés döntő mértékben függ az alkalmazott energiafüggvényektől. Ezért aktívan kutattunk új módszereket hatékony statisztikai potenciálok számítására. Egy új statisztikai párpotenciált fejlesztettünk ki, amely az energiafüggvény referencia-állapotának egy új definícióját használja.

2.5. Molekulamodellezés biológiai alkalmazásai

Módszereinket számtalan együttműködésben hasznosítottuk kísérletes csoportokkal, ezek közül mutatunk be néhányat enzim specificitás tervezése, enzim mechanisztikai vizsgálatok, fold verifikáció, gyógyszer-indukált konformációváltozások feltérképezése és cirkulárisan permutált molekulaserkezet modellezése témakörökben.

3. EREDMÉNYEK ÉS DISZKUSSZIÓJUK

3.1. Módszerek általános összehasonlító molekulaszerkezet-modellezésre

3.1.1 Multiple Mapping Method – egy új szerkezet alapú szekvencia-illesztési módszer

Rai BK, Fiser A. *Proteins* (2006) 63(3): 644-61;

Rai BK, Madrid-Aliste CJ, Fajardo JE, Fiser A. *Bioinformatics* (2006) 22(21) : 2691-2;

A szekvencia-illesztés az egyik legkritikusabb pontja a homológián alapuló térszerkezet modellezésnek. Egyetlen aminosav helytelen szekvencia-illesztése kb. 4 Å torzulási hibát okoz a szerkezetben (átlagos hossza a C_{alfa}-C_{alfa} távolságnak). A szekvencia-illesztési eljárások gyakran alapulnak dinamikus programozási algoritmuson, ami garantálja az energiaminimum megtalálását az adott rendszerben. Nehéz szekvencia-illesztések esetén (kevesebb mint 20-30%-os globális szekvencia azonosság a fehérjék között) a különböző szekvencia-illesztési eljárások egyes szegmenseket eltérő módon illeszthetnek. Tekintettel arra, hogy minden alternatív megoldás az adott módszer szerint optimális, egy információgazdagabb rendszerben lehet csak kiértékelni, hogy melyik illesztés a helyesebb. A homológia modellezés esetén a szekvencia-illesztés egy sajátos tulajdonsággal rendelkezik – az illesztés egyik oldalán nem csak a fehérje szekvenciája, hanem szerkezete is ismert. Következésképpen az illesztés pontosságát úgy lehet növelni, hogy térszerkezeti információt is figyelembe veszünk.

Az általunk kidolgozott új eljárás, a Multiple Mapping Method (MMM) közvetlenül hasznosítja a háromdimenziós szerkezeti információt. A módszer oly módon minimalizálja az illesztési hibák előfordulását, hogy az alternatív szekvencia-illesztések megoldásainak eltérően illesztett szegmenseit kiértékeli a térszerkezet ismeretében. Ezután a pontosan illesztett szegmenseket kombinálja egy végső, optimális megoldáshoz. Az illesztések kiválasztását egy energiafüggvény vezérli, amely kiértékeli minden egyes alternatívan illesztett szegmens kompatibilitását a templát szerkezeti környezetében. Az energiafüggvény négy komponensből áll: (1) környezetfüggő mutációs mátrixból (2) aminosav mutációs mátrixból (3) H3P2 mátrixból, ami a modellezett szekvenciára jósolt és a templát megfelelő szegmensében megfigyelt másodlagos szerkezeti elemek illeszkedését értékeli ki és (4) statisztikai potenciálból, ami a nem-kovalens kölcsönhatásokat számolja a szegmens és környezete között (ennek kifejtéséről szól bővebben a 3.4. fejezet). Az energiafüggvény egymástól eltérő típusú paramétereit dimenziómentes statisztikai Z-értékekké konvertáljuk és így azok egy

homogén rendszert alkotnak. A Z-értékes megoldás további előnye, hogy azok a paraméterek, amik nem szignifikánsak egy adott probléma szempontjából automatikusan kisebb súllyal szerepelnek a kiértékelésben. Ennek az a biológiai értelme, hogy egy szekvencia-illesztés esetén néha szekvenciajel, néha másodlagos szerkezeti preferenciák, megint máskor térszerkezeti szükségszerűségek (kontaktusok) vezérelhetik az illesztést, és ez ugyanabban a szekvenciában szegmensenként változhat. Az MMM eljárás gyakorlatilag a szekvencia-illesztést optimalizálja azáltal, hogy a kritikus, nehezen illeszthető szegmenseket minden biológiailag értelmezhető módon ráilleszti a templátra és a templát szerkezeti környezetében kiértékeli az illesztések alkalmasságát. A módszernek két előnye van a hagyományos szekvencia alapú illesztőeljárásokkal szemben, és mindkét előny a templátszerkezettel hozható összefüggésbe. Egyrészt minden aminosavpár illesztéséhez megkeresi a szekvenciálisan távoli, de térben közeli aminosavakat, amik a térbeli környezetet alkotják, így módon minden egyedi aminosavpár illesztés problémáját megkönnyíti, mert többlet információt használ fel a térbeli közelségben levő további aminosavak illesztésének vizsgálatával. Másrészt a templátszerkezet lehetőséget ad arra, hogy a hagyományos aminosav-kicserélődési mátrixon túlmenően egy sokkal teljesebb energiafüggvénnyel történjen az illesztés, ami figyelembe veszi a másodlagos szerkezeti motívumokat, kiértékeli a 3D kontaktusokat, és olyan aminosav kicserélődési mátrixot tud alkalmazni, ami környezetfüggő, azaz tekintettel van az aminosavak oldószer exponáltságára, hidrogén-kötésmintázatára, másodlagos szerkezetére stb. is.

Tesztelések során az MMM módszer mindkét másik modellező eljárásnál, ami hasonló céllal készült, Esypred /Lambert és mtsi. *Bioinformatics* 2001, 9, 1250/ és Consensus /Prasad és mtsi. *Bioinformatics* 2003, 19, 1682/ pontosabbnak bizonyult. A szakmai közösség számára egy szerver formájában hozzáférhető az eljárás, amit az elmúlt 2 évben több mint 3000 alkalommal használtak.

3.1.2. M4T – Homológia modellezés többszörös templáttal

Fernandez-Fuentes N, Rai BK, Madrid-Aliste CJ, Fajardo JE, Fiser A. *Bioinformatics* (2007) 23(19): 2558-65

Fernandez-Fuentes N, Madrid-Aliste CJ, Rai BK, Fajardo JE, Fiser A. *Nucleic Acids Res* (2007) 35: W363-8

Az MMM módszert később kibővítettük egy teljes, automatikus modellező eljárássá. Ehhez arra volt szükség, hogy a korábbi manuális templát kiválasztás helyett egy algoritmikus megoldást adjunk. Annak érdekében, hogy a homológia modellezés teljes lehetőségét kihasználjuk, kidolgoztunk egy iteratív klaszterező eljárást, ami több templátszerkezetet képes azonosítani és optimálisan kombinálni. A templátok kombinálása két előnnyel is járhat.

Egyrészt a templátok átfedő részein pontosabb modellt lehet kiszámítani, másrészt a nem átfedő részek segítségével a modell méretét lehet növelni. Az iteratív klaszterező eljárás első lépése a lehetséges templátok azonosítása, amit szekvencia-profillal indított heurisztikus szekvencia-illesztéssel kapunk meg (sequence profile initiated PSI-Blast alignment). A szekvenciaprofilokat a BlastProfiler program segítségével készítjük elő, amit korábban a csoportunkban dolgoztunk ki. A lehetséges templátokat egy hierarchikus klaszterező folyamat során választjuk ki, amely figyelembe veszi a templátok egymással vett szekvencia azonosságát, a templátok és a modellezett szekvencia azonosságát, a templátok kísérletileg megoldott kristályszerkezetének a felbontását, a doménszerkezet lefedettségét és az egyes templátok egyedi hozzájárulását a modellezett szekvenciához. Ez utóbbi azt jelenti, hogy ha egy templát először nem kerül kiválasztásra, mert például egy másik templát jobban hasonlít a modellezett szekvenciára, vagy mert a másik templát kísérleti szerkezete jobb felbontású, utólag még mindig felhasználásra kerülhet, ha van egy olyan szegmense, ami hiányzik a többi szerkezetből. A fenti munkák gyakorlati eredménye az M4T modellező programcsomag (Multiple Mapping Method with Multiple Templates), ami egy online szervernek köszönhetően bárki számára elérhető és amit mi magunk is használtunk kutatási együttműködéseinkben (3.5 fejezet). A CASP6 konferencia tesztmodelljein az M4T a második legjobb teljesítményt nyújtotta, beleértve olyan csoportokat is, akik manuálisan szerkesztették modelljeiket.

3.2. Loopok modellezése

Az M4T modellező programcsomag, és általában minden homológia szerkezetmodellező eljárás, a szekvencia azon részére tud modellt építeni, amihez van templát információ. Gyakori kivétel ez alól a legtöbbször a molekula felszínén elhelyezkedő loopok esete. A loopok funkcionális szerepe lehet antitest felismerés, ligandkötő helyek kialakítása, például kalcium, NAD(P), DNS számára, vagy az enzimek aktív centrumának létrehozása, például proteázoknál. Mivel ezek gyakran felelősek a fehérje specificitásáért, ezért nem meglepő, hogy a templát és a modellezett szekvencia között a leggyakoribb különbségek is ezekben a régiókban lépnek fel. A templátban vagy nem létezik hasonló szegmens vagy annyira eltérő, hogy nem alkalmas modellezésre. A loop szegmensek modelljeinek pontossága kritikus jelentőséggel bír az egész molekulamodell alkalmazhatósága szempontjából, ha ligandum-dokkolás vagy funkcionális jellemzés a feladat. A loop- modellezés általában egy önálló feladat, ami a fehérje-modellezést követi, amikor is a hiányzó vagy pontatlanul modellezett loop szegmenseket hozzáadjuk a modellhez. A loop-modellezést fel lehet fogni

egy mini fehérje folding problémának, azzal a fontos különbséggel, hogy ezek a szegmensek általában túl rövidek ahhoz, hogy a lokális foldról elégséges információt szolgáltassanak. Továbbá a loopok fiziológiás környezete nem pusztán az oldószer, mint az egész fehérjelánc felcsavarodása esetén, hanem maga a loop kezdő és végpontjait horgonyzó fehérjemolekula. Egy-két ritka esetben decapeptideket is megfigyeltünk teljesen eltérő konformációban (kaméleon szekvenciák), ha az eltérő fehérjeszerkezeti környezet különböző konformációba kényszerítette ezeket a szegmenseket.

Hasonlóan az egész fehérje-modellezéséhez, a loop-modellező eljárásoknak két fő irányvonala van (1) fragmenskereső módszerek, amelyek a homológia fehérje-modellezéshez hasonlóan arra irányulnak, hogy egy hasonló konformációjú szegmenst találjanak bármelyik ismert fehérje szerkezetben, és (2) konformáció optimalizáló eljárások, amelyek az *ab initio* megközelítéshez hasonlóak.

3.2.1. Fragmens kereséses loop-modellezés

Fernandez-Fuentes N, Zhai J, Fiser A. *Nucleic Acids Res* (2006) 34: W173-6

Fernandez-Fuentes N, Fiser A. *BMC Struct Biol* (2006) 6, 15

Fernandez-Fuentes N, Oliva B, Fiser A. *Nucleic Acids Res* (2006) 34(7): 2085-97

A fehérjeszerkezet univerzum sok részletére derült fény az elmúlt évtized során a szerkezetgenomikai kutatásoknak köszönhetően. Jelentősen bővült a rövid szegmensek konformációs repertoárja is. Saját kutatásunk szerint ma már szinte az összes a természetben megvalósuló 10 aminosavnál rövidebb szegmenshez lehet találni legalább egy ismert konformációt, amivel legalább 50%-os szekvencia azonosságot mutat. Azt is kimutattuk, hogy habár minél rövidebb szekvenciákat hasonlítunk össze, annál magasabb szekvencia azonosság kell a szerkezeti hasonlóság feltételezéséhez, 50%-os azonosság még egész rövid szegmensek esetén (4-14 aminosav) is biztosítja a szerkezeti hasonlóságot. Továbbá kimutattuk, hogy annak ellenére, hogy 2002 után 5 éven belül a Fehérjeszerkezeti Adatbank mérete megduplázódott, a szekvencia adatbázisok pedig meghatszorozódtak, egyetlen olyan új szerkezeti szegmens sem került elő, ami ne hasonlítana már ismert konformációkra 10 aminosavnyi szegmenshosszig, és nem került elő olyan szekvenciaszegmens, ami ne hasonlítana legalább 50%-ban egy ismert konformációhoz. Ez arra utal, hogy a fehérjeszerkezetek gyakran újrahaznosítanak már meglévő rövid építőelemeket. Ez a felismerés új lökést adott a fragmens alapú loop-modellező eljárásoknak. Az általunk kidolgozott eljárás (ArchPred) a következő módon működik. Az eljárás alapja egy hierarchikus loop fragmens adatbázis, amiben jelenleg kb. 300 000 fragmens reprezentálja az összes ismert, természetben megvalósuló egyedi loop konformációt. A loop

fragmenseket hosszuk és az őket övező másodlagos szerkezetek típusán kívül a konformációkat leíró négy belső koordináta szerint jellemeztük, melyek három diéderes szögből és egy távolságból állnak össze. A megfelelő konformációk ebből a fragmens adatbázisból kerülnek kiválasztásra aszerint, hogy a loop hossza, a környező másodlagos szerkezetek és a négy belső koordináta értékei illeszkedjenek a modellben megfigyelttekhez. A kiválasztott szegmenseket beillesztjük a fehérje szerkezetébe, ahol a szegmens és a csatlakozópontjai RMSD illeszkedését vizsgáljuk, továbbá, hogy a kiválasztott konformáció mennyi térbeli ütközés árán illeszkedik a szerkezeti környezetbe. Az utolsó lépésben a megmaradt fragmenseket aszerint rangsoroljuk, hogy szekvenciálisan mennyire hasonlók a modellezett loophoz, és hogy a peptidkötésnek a konformációkban megfigyelt és a szekvenciából kiszámolt ϕ - ψ diéderes szögei mennyire illeszkednek. A legmegfelelőbbnek talált konformációt beillesztjük a szerkezetbe és oldallánc rekonstrukció után egy rövid konjugált gradiens energiaminimalizációval „beolvasztjuk” a szerkezetbe. Az egész protokoll teljesen automatizált, és egy web szerveren keresztül elérhető.

Munkánk során azt a konklúziót vonhattuk le, hogy a rövid szegmensek kimerítő reprezentációja az ismert fehérjékben arra utal, hogy a fragmenskereső loop-modellező módszerek inkább a megfelelő konformáció felismerésére kell, hogy fókuszáljanak, mert ez teszi a módszerek alkalmazását hatékonyá, és nem kell tartani a mintavétel szegényességétől, mint azt egyes korábbi munkák állították. Ennek megfelelően az ArchPred módszerünk kidolgozásánál igen nagy hangsúlyt fektettünk a predikció energiafüggvényének optimalizálására.

3.2.2. Konformáció kereséses loop-modellezés

Fiser A, Do RK, Sali A. *Protein Sci* (2000) 9(9) : 1753-73

Fiser A, Feig M, Brooks CL 3rd, Sali A. *Acc Chem Res* (2002) 35(6) : 413-21

Fiser A, Sali A. *Bioinformatics* (2003) 19(18) : 2500-1

A fragmenskereső eljárások nem minden esetben képesek a megfelelő konformációt azonosítani loop-modellezésben. Ennek a technikai korlátozásnak a kiküszöbölésére *ab initio* alapú, konformáció kereső eljárást fejlesztettünk ki. A ModLoop nevű modellező eljárásunk konjugált gradiens minimalizációt kombinál molekula dinamikával, amit szimulált hőkezeléses protokollba ágyaztunk. Létrehoztunk egy pszeudo-energiafüggvényt, ami több különböző típusú energiafüggvény kombinációja. Ezek részben a CHARMM molekula mechanikai erőkből származnak, részben pedig statisztikai „egy-test” és „két-test” potenciálok. Az előbbit peptidkötések diéderes szögeinek statisztikus előfordulási valószínűségéből számoltuk ki, az utóbbi pedig egy aminosav párpotenciál, amit a kölcsönható aminosavak típusa és a köztük

levő távolságok előfordulásából számoltunk ki az inverz Boltzmann összefüggést használva. Annak érdekében, hogy a modellező eljárás gyakorlati problémákra jól alkalmazható legyen, az energiafüggvény és az optimalizációs protokollt 400 különböző eseten teszteltük, és mind natív, mind fokozatosan eltorzított környezetben is vizsgáltuk a hatékonyságát. Ez utóbbira azért volt szükség, mert ezeket a loopokat legtöbbször molekulamodellekbe kell beépíteni, ahol a loopok szerkezeti környezete általában csak hozzávetőlegesen pontos. Ugyanakkor röntgenkrisztallográfiában is gyakran alkalmazzák a módszerünket, ahol a diffrakciós kép hiányossága gyakran függ össze felszíni loopok konformációjával. A módszerünk pontosságát 8 aminosavnyi hosszú loopokra szemléltetem. Amennyiben az <1 , $1-2$ és >2 Å RMSD eredményeket jónak, elfogadhatónak illetve rossznak osztályozzuk, akkor hasonló sorrendben, a ModLoop eljárás esetén a tesztelt modellek 50, 40, és 10%-a esik ezekbe a kategóriákba.

Az összehasonlító vagy homológia molekulamodellezési eljárások egyik egyszerűsítése, hogy az erőterek alkalmazása vákuumban történik. Ezt az egyszerűsítést egyrészt a modellek pontossága utólag igazolja, másrészt pedig az oldószerhatás számítása rendkívül költséges (időigényes), ezért technikailag ritkán szokták kivitelezni. Loopok esetében viszont a modellezett rendszer elég kicsi ahhoz, hogy az oldószerhatást figyelembe vehessük. A nemrég kifejlesztett implicit Általánosított Boltzmann oldószerhatást alkalmaztuk a ModLoop-ban. Az oldószerhatás figyelembevétele egy nagyságrenddel hosszabbította meg a modellezés időigényét, ugyanakkor a modellek minősége - ha nem is drámai mértékben, de - tovább javult. Ugyanakkor, ha a vákuumban végzett számításainkat az eredetnél 10-szer hosszabb ideig optimalizáltuk, akkor hasonló, további kismértékű javulást értünk el. Összefoglalva, amíg biológiai-fizikai szempontból nem kétséges, hogy az oldószerhatást figyelembe kell venni, hogy valós képet kapjunk egy fehérjeszerkezetről annak fiziológiás környezetében, de a jelenlegi erőterek nem elég pontosak ahhoz, hogy ez a modellek pontosságában számszerűen is jelentkezzen. A ModLoop eljárás elérhető mint önálló szerver, és mint a Modeller térszerkezet modellező csomag része is.

3.3. Eljárás pontmutációk modellezése

Feyfant E, Sali A, Fiser A. *Protein Sci* (2007) **16**(9) : 2030-41

Pontmutációk létrehozása a fehérjeszerkezetben az egyik leggyakoribb eljárás a funkció tesztelésére, a folding folyamatának, vagy a szerkezet stabilitásának vizsgálatára. Pontmutációk segítségével lehet nehézatomban kötőhelyeket bevezetni a szerkezetbe röntgenkrisztallográfiái kísérletekhez, vagy más spektrális jeladókat, például spin jeleket. Egy teljes fehérjemolekula

modellezésével szemben, a pontmutációk esetén kihívás, hogy csak az igen nagy pontosságú modelleknek van haszna, hiszen egy alig pár tízed angstromos torzulás egy aktív centrumban található aminosav esetén a funkció teljes elvesztéséhez vezethet.

Az általunk kidolgozott pontmutáció-modellezés (Mutate_Model eljárás) esetében, a korábbi munkáinkhoz hasonlóan nagy hangsúlyt fektettünk az energiafüggvény optimalizációjára, mert a modellek pontosságának javulásában tapasztalatunk szerint ez a fő korlátozó tényező. Ez különösen igaz lehet pontmutációk modellezése esetén, ahol ún. oldallánc konformerek – az oldalláncok sztereo-kémiai preferált konformációs állapotai – nagymértékben leegyszerűsítik a konformációs teret. Korábbi módszerekkel ellentétben az általunk fejlesztett energiafüggvénynél nem a konformerek statisztikai preferenciáját használtuk, hanem kombináltuk a molekulamechanikai erőterek használatát az ún. „homológiai kényszerítő erőkkel”. Ezek a kényszerítő erők az adott oldallánc preferált konformációiból és a konkrétan kicserélésre szánt oldallánc adott konformációjából származnak. Az adott aminosav megfigyelt konformációinak preferált eloszlására a legkisebb négyzetek módszerével egy folytonos függvényt illesztettünk, amiből normalizálás után az adott konformációnak a valószínűségi sűrűségfüggvényét kaptuk. Az egyes konformációkhoz tartozó valószínűségi sűrűségfüggvényeket közvetlenül tudtuk alkalmazni az összetett energiafüggvényünkben. Az energiafüggvénynek 36 kombinációját vizsgáltuk 717 szerkezetileg ismert pontmutáció modellezésén keresztül. A homológiai kényszerítő erőkből az oldalláncok diédes szögeire vonatkozó bizonyult a leghasznosabbnak (mind statisztikai megfigyelésekből, mind a templátszerkezetből), míg a távolságra vonatkozó kényszerítő erők inkább rontották az eredményeket. Ez nagy valószínűséggel annak köszönhető, hogy ezek túlságosan leszűkítették és a vad típusához hasonlóvá tették az oldallánc lehetséges konformációját. A Mutate_Model eljárás a Modeller csomag részeként elérhető a szakmai közönség számára.

3.4. Statisztikai párpotenciál kifejlesztése

Rykunov D, Fiser A. *Proteins* (2007) 67(3) : 559-68

Az összes eddig ismertetett modellező eljárás – beleértve a szekvencia-illesztést is – felfogható optimalizációs problémának. Optimalizáció során három fő tényezőt kell tisztázni és egymással harmonizálni: (1) a rendszer definícióját (például az atomok három dimenziós koordinátái fehérjében, vagy az aminosavak szekvenciája stb.) (2) a kereső algoritmust, amivel feltérképezzük a rendszert (pl. koordináták esetén molekula dinamikai szimuláció vagy szekvencia-illesztés esetén dinamikus programozás), (3) és az energiafüggvényt, amivel

jellemezzük a rendszer állapotait. A megfelelő energiafüggvény kritikus eleme az optimalizációnak, hiszen hiába van egy tökéletes kereső algoritmusunk, ha nem tudjuk a rendszer optimális állapotait felismerni. Munkám egyik fontos eleme új energiafüggvények előállítása volt. A fehérjéket alkotó atomok nagy száma és a kölcsönhatások komplexitása miatt empirikus statisztikai párpotenciálokat fejlesztettünk, mind atomi, mind aminosav szinten. Ezen potenciálok alkalmazhatósága szorosan összefügg a statisztikai megfigyelések pontosságával. Az ún. kvázi-kémiai megközelítés egy gyakorta alkalmazott elméleti alap a fehérjeszerkezet alapú párpotenciálok kifejlesztéséhez. Ez lehetőséget ad a statisztikus mechanika elveinek alkalmazására. A kvázi-kémiai megközelítés feltételezi, hogy a rendszer független elemekből áll és csak párkölcsönhatások működnek benne. Fehérjék esetén ez egy elfogadható megközelítés kontaktpotenciálokra, de erősen sérül távolságfüggő potenciáloknál, ahol a nagyobb távolságok akár a fehérje méretével is összevethetőek. Kutatásunk során azt vizsgáltuk, hogy milyen eredetű és mekkora mértékű a kvázi-kémiai megközelítéssel levezetett statisztikai potenciálok hibája, amikor fehérjeszerkezetekre alkalmazzuk. A statisztikai potenciálok távolságfüggését randomizált fehérjeszerű modelleken elemeztük. Ilyen modellekből levezetett potenciálok átlagos értéke nulla kell, hogy legyen minden vizsgált távolságnál. Megfigyelésünk szerint, ez a „referencia-potenciál” szisztematikusan nullától jelentős mértékben eltérő értékeket vesz fel. Kutatásunk során rámutattunk, hogy ez a hamis pozitív energiatartalom két dologgal függ össze: (1) rövid távolságokon belül a korlátozott számú statisztikai megfigyeléssel és (2) nagyobb távolságoknál a fehérjék korlátozott méretével. Megállapítottuk, hogy ezek a szisztematikus hibák összefüggésbe hozhatóak a fehérjék méretfüggő aminosav összetételével és változó méretével. Számításaink szerint, 4 Å-nél rövidebb távolságoknál alig 50%-a szignifikáns a statisztikai megfigyeléseknek, és még hosszabb távolságoknál is csak maximum 80%-a, ami döntő hatással van az ezekből számított energiafüggvényekre.

Ezen megfigyeléseink alapján módosítottuk a statisztikai potenciálok referencia-állapotának definícióját, és a korábbi eljárásokkal szemben, amik a fehérjék különböző környezetét átlagolták ki, mi fehérjeszerű, randomizált modelleket használtunk, amelyek megőrizték a mintául szolgáló fehérjék összetételét, méretét és sűrűségét, de semmilyen kötési rendet nem tartottak meg. Ezzel a referencia-állapot definícióval sikerült a hamis pozitív jelet megszüntetni a megfigyelt kölcsönhatási statisztikákban és javítani az ezekből levezetett potenciálok hatékonyságát. Ezt az újonnan kifejlesztett statisztikai párpotenciált használtuk fel a szekvencia-illesztő MMM eljárásban, hogy a fragmensek kölcsönhatását kiértékeljük az adott fehérjekörnyezetben, és ezt a potenciált használtuk hasonló célra az ArchPred fragmens kereső

loop modellező programban is.

3.5. Molekulamodellező módszerek gyakorlati alkalmazásai

3.5.1. Enzim specificitás tervezése

Wu, G., Fiser, A., ter Kuile, B., Sali, A. and Muller, M. (1999) Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. *Proc.Natl.Acad.Sci.U.S.A.*, **96**, 6285.

A *Trichomonas vaginalis* genomjában egy olyan laktát dehidrogenázt találtunk, ami nagyobb szekvenciahasonlóságot mutatott a saját malát dehidrogenázához, mint bármilyen más szervezetben fellelhető laktát dehidrogenázhoz. Filogenetikai vizsgálatok arra utaltak, hogy a *T. vaginalis* laktát dehidrogenáza a viszonylagos közelmúltban konvergensen evolválódott a malát dehidrogenázokból. Molekulamodellek építésével és vizsgálatával lokalizáltuk azokat az aminosavakat, amik az új enzimspecificitásban szerepet játszhattak. A molekulaszervezetből nyert információval célzott mutagenézis-vizsgálatokat végeztünk, amik sikeresen konvertálták a *T. vaginalis* laktát dehidrogenázát malátspecifikus enzimmé.

3.5.2. Enzim mechanisztikai vizsgálatok

Fiser, A. and Vertessy, B.G. (2000) Altered subunit communication in subfamilies of trimeric dUTPases. *Biochem.Biophys.Res.Comm.*, 279, 534.

A dUTP-áz enzimek megakadályozzák az uracil beépülését a DNS-be, így ha a faj-specifikus enzimmechanisztikus tulajdonságait lehetne azonosítani ezekben a szervezet számára létfontosságú molekulákban, akkor fontos gyógyszer célpontokká válhatnak. A dUTP-ázok általában homotrimer szerkezetű enzimek, ahol a három azonos aktív centrum öt különböző szekvenciamotívumból épül fel oly módon, hogy mindegyik aktív centrum mind a három alegységből kap motívumokat. A humán, *E. coli*, *D. melangosterből*, DNS- és retrovirusokból szekvenált dUTP-áz molekulák térszerkezeti modelljeiben térképeztük fel a konzervált aminosav-mintázatokat. Megfigyeltük, hogy az evolúció során az alegység-kommunikáció ezekben a homotrimer enzimekben két különböző típusú alcsaládra választja szét a fehérjét. A magasabb rendű élőlényekben poláros és töltött aminosavak játszanak szerepet az alegységek közötti kommunikációban – feltehetően az allosztérikus kontroll domináns megjelenése miatt –, míg baktériumoknál és vírusoknál az alegységek kölcsönható felszínei még teljesen apoláros jellegűek.

3.5.3. Fold verifikáció

Wu, G., McArthur, A.G., Fiser, A., Sali, A., Sogin, M.L. and Müller, M. (2000) *Mol.Biol.Evol.*, 17, 1156.

Hiszton molekulákat nem találtak baktériumokban, és sok archae-ban sem. Hisztonszerű molekulákat viszont megfigyeltek az euryarchaeotákban, de ezek markánsan különböznek az eukarióta hisztonoktól. A *Giardia lamblia* egy mitokondrium nélküli egysejtű eukarióta, amit filogenetikai vizsgálatok az egyik legelső eukarióta leágazásának tekintenek és emiatt sokáig úgy gondolták, hogy alapvető tulajdonságait a legősibb maggal rendelkező sejtekből őrizte meg, így például nem rendelkezik hisztonokkal. A *Giardia lamblia* frissen szekvenált genomjában találtunk négy olyan szekvenciát, amik hasonlóságot mutattak eukarióta hisztonokkal. Megépítettük a molekulamodelljeiket a béka ismert hiszton szerkezetei alapján, majd energia-függvények számításával teszteltük a molekula-szerkezetek stabilitását. A tesztjeink során kiderült, hogy az energiaszámítások elég érzékenyek ahhoz, hogy különbséget tudjanak tenni a béka különböző saját hisztonjai között is. Ha például a H2a hiszton szekvenciáját a H2b szerkezet alapján modelleztük, akkor kimutatható instabilitást észleltünk. Ezután kiszámoltuk a *G. lamblia* molekulaszerkezeteinek az energiastabilitását és azt találtuk, hogy azok a békához hasonlóan stabilak. Ebből azt a következtetést vontuk le, hogy a *G. lamblia* nem képvisel átmeneti szerveződést az archae és az eukarióták között, hanem egy tipikus eukarióta szerveződés. Továbbá, hogy a hiszton alapú kromatin szerveződés már megjelent a legelső eukariótáknál és csak nagyon kismértékű változásokon esett át az azóta eltelt időben.

3.5.4. Gyógyszer-indukált konformációváltozások feltérképezése tubulinban

Xiao, H., Verdier-Pinard, P., Fernandez-Fuentes, N., Burd, B., Angeletti, R., Fiser, A., Horwitz, S.B. and Orr, G.A. (2006) *Proc. Natl. Acad. Sci. USA*, 103, 10166-10173.

A Taxol nevű tumorellenes gyógyszermolekula a mikrotubulinokhoz kapcsolódva stabilizálja azokat, korlátozza azok dinamikáját, sejthalált mozdítva elő. A tubulin filamentumokban Taxol hatására végbemenő szerkezeti változásokat tanulmányoztuk. Hidrogén/deutérium kicserélődéshez kapcsolt HPLC elektropray ionizációs tömegspektrometriai kísérleteket végeztünk, mind a natív mikrotubulinban, mind amikor Taxol kapcsolódott hozzá. A molekulamodellzési feladat az volt, hogy az alfa és béta tubulin alegységek filamentális negyedleges szerkezetét létrehozzuk és a szerkezetben feltérképezzük a dinamikai változásokat. A natív protofilament alfa és béta tubulin dimerek fej-láb kapcsolatok

révén alakulnak ki, és alkotnak aztán további laterális kapcsolatokat. Míg a tubulinok polimer szerkezete csak cryo-elektromikroszkópos felvételtől ismert 8Å felbontásban, a dimer szerkezetét sikerült röntgenkristallográfiával megoldani egy hasonló genomból. A modellezés részben az alegységek atomi modelljeit építette meg, részben docking eljárással az egész filamentumot rekonstruálta. Ezen szerkezeti modellen vizsgáltuk a kísérletes adatokat, amik konformációváltozásra utaltak és ennek révén új dimer-dimer kölcsönhatásokat sikerült azonosítani.

3.5.5. Cirkulárisan permutált molekulaszerkezet modellezése

Barrientos, L.G., Campos-Olivas, R., Louis, J.M., Fiser, A., Sali, A. and Gronenborn, A.M. (2001) *J.Biomol.NMR*, 19, 289.

A Cyanovirin-N fehérjét (CV-N) cianobaktériumból izolálták, és megfigyelték, hogy a HIV-t burkoló cukorfehérjékhez kapcsolódik és ezáltal nélkülözhetetlen biokémiai kölcsönhatásokat gátol, amik a vírus és a gazdasejt fúzióját akadályozzák meg. Emiatt a CV-N a vizsgálatok középpontjába került mint potenciális gyógyszer analóg. Szerkezeti vizsgálatok spontán domén kicserélődést figyeltek meg a CV-N szerkezetekben. Ennek részletesebb vizsgálatára cirkulárisan permutált fehérjét állítottak elő. Ennek a szerkezetét nem oldották meg, viszont NMR segítségével ^1H , ^{13}C és ^{15}N rezonancia spektrum asszignációt hajtottak végre. Az általunk épített modell azzal a feltételezéssel élt, hogy a domének belső szerkezete nem változott, ezért a modellezés gyakorlatilag a modelleket összekötő loop-modellezésére egyszerűsödött, amit a ModLoop eljárással hajtottuk végre. Az épített modellből számolt elméleti rezonancia hozzárendelések kitűnő egyezést mutattak a kísérletivel, aminek segítségével kimutattuk, hogy a cirkulárisan permutált CN-V általános szerkezete megegyezik a vad típusával.

4. PUBLIKÁCIÓK

4.1. A Ph.D. fokozat megszerzése óta megjelent és az értekezésben felhasznált publikációk listája

1. Rai, B.K., Madrid-Aliste, C.J., Fajardo, J.E. and Fiser, A. (2007) MMM: a sequence-to-structure alignment protocol. *Bioinformatics*, **22**, 2691-2692.
2. Rai, B.K. and Fiser, A. (2006) Multiple mapping method: a novel approach to the sequence-to-structure alignment problem in comparative protein structure modeling. *Proteins*, **63**, 644-661.

3. Fernandez-Fuentes, N., Rai, B.K., Madrid-Aliste, C.J., Fajardo, J.E. and Fiser, A. (2007) Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics*, **23**, 2558-2565.
4. Fernandez-Fuentes, N., Madrid-Aliste, C.J., Rai, B.K., Fajardo, J.E. and Fiser, A. (2007) M4T: a comparative protein structure modeling server. *Nucleic Acids Res*, **35**, W363-368.
5. Fiser, A., Feig, M., Brooks, C.L., III and Sali, A. (2002) Evolution and Physics in Comparative Protein Structure Modeling. *Acc.Chem.Res.*, **35**, 413.
6. Fiser, A., Do, R.K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.*, **9**, 1753.
7. Fiser, A. and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics.*, **19**, 2500.
8. Fernandez-Fuentes, N., Zhai, J. and Fiser, A. (2006) ArchPRED: a template based loop structure prediction server. *Nucleic Acids Res*, **34**, W173-176.
9. Fernandez-Fuentes, N. and Fiser, A. (2006) Saturating representation of loop conformational fragments in structure databanks. *BMC Struct Biol*, **6**, 15.
10. Fernandez-Fuentes, N., Oliva, B. and Fiser, A. (2006) A supersecondary structure library and search algorithm for modeling loops in protein structures. *Nucleic Acids Res*, **34**, 2085-2097.
11. Feyfant, E., Sali, A. and Fiser, A. (2007) Modeling mutations in protein structures. *Protein Sci*, **16**, 2030-2041.
12. Rykunov, D. and Fiser, A. (2007) Effects of amino acid composition, finite size of proteins, and sparse statistics on distance-dependent statistical pair potentials. *Proteins*, **67**, 559-568.
13. Wu, G., Fiser, A., ter Kuile, B., Sali, A. and Muller, M. (1999) Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. *Proc.Natl.Acad.Sci.U.S.A*, **96**, 6285.
14. Fiser, A. and Vertessy, B.G. (2000) Altered subunit communication in subfamilies of trimeric dUTPases. *Biochem.Biophys.Res.Comm.*, **279**, 534.
15. Wu, G., McArthur, A.G., Fiser, A., Sali, A., Sogin, M.L. and Mllerm, M. (2000) Core histones of the amitochondriate protist, *Giardia lamblia*. *Mol.Biol.Evol.*, **17**, 1156.
16. Xiao, H., Verdier-Pinard, P., Fernandez-Fuentes, N., Burd, B., Angeletti, R., Fiser, A., Horwitz, S.B. and Orr, G.A. (2006) Insights into the mechanism of microtubule stabilization by Taxol. *Proc Natl Acad Sci U S A*, **103**, 10166-10173.
17. Barrientos, L.G., Campos-Olivas, R., Louis, J.M., Fiser, A., Sali, A. and Gronenborn,

A.M. (2001) ^1H , ^{13}C , ^{15}N resonance assignments and fold verification of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. *J.Biomol.NMR*, **19**, 289.

4.2. A Ph.D. fokozat megszerzése óta megjelent és az értekezésben fel nem használt publikációk listája

18. Manjasetty, B.A., Shi, W., Zhan, C., Fiser, A. and Chance, M.R. (2007) A high-throughput approach to protein structure analysis. *Genet Eng (N Y)*, **28**, 105-128.
19. Bonanno, J.B., Almo, S.C., Bresnick, A., Chance, M.R., Fiser, A., Swaminathan, S., Jiang, J., Studier, F.W., Shapiro, L., Lima, C.D. *et al.* (2005) New York-Structural GenomiX Research Consortium (NYSGXRC): A Large Scale Center for the Protein Structure Initiative. *J Struct Funct Genomics*, **6**, 225-232.
20. Chance, M.R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J.E., Radhakannan, T. and Marinkovic, N. (2004) High-throughput computational and experimental techniques in structural genomics. *Genome Res.*, **14**, 2145.
21. Pieper, U., Eswar, N., Braberg, H., Madhusudhan, M.S., Davis, F.P., Stuart, A.C., Mirkovic, N., Rossi, A., Marti-Renom, M.A., Fiser, A. *et al.* (2004) MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.*, **32**, D217.
22. Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V.A., Pieper, U., Stuart, A.C., Marti-Renom, M.A., Madhusudhan, M.S., Yerkovich, B. *et al.* (2003) Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.*, **31**, 3375.
23. Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461.
24. Fiser, A., Sanchez, R., Melo, F., Sali, A., Watanabe, M., Roux, B., MacKerell, A.D., Jr. and Becker, O. (2001), *Computational Biochemistry and Biophysics*. Marcel Dekker, pp. 275.
25. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu.Rev.Biophys.Biomol.Struct.*, **29**, 291.
26. Torres, M., Fernandez-Fuentes, N., Fiser, A. and Casadevall, A. (2007) Exchanging murine and human immunoglobulin constant chains affects the kinetics and thermodynamics of antigen binding and chimeric antibody autoreactivity. *PLoS ONE*, **2**, e1310.
27. Almo, S.C., Bonanno, J.B., Sauder, J.M., Emtage, S., Diloranzo, T.P., Malashkevich, V., Wasserman, S.R., Swaminathan, S., Eswaramoorthy, S., Agarwal, R. *et al.* (2007) Structural genomics of protein phosphatases. *J Struct Funct Genomics*. **8** (2-3), 121-140

28. Torres, M., Fernandez-Fuentes, N., Fiser, A. and Casadevall, A. (2007) The immunoglobulin heavy chain constant region affects kinetic and thermodynamic parameters of antibody variable region interactions with antigen. *J Biol Chem*, **282**, 13917-13927.
29. Reichman, C., Singh, K., Liu, Y., Singh, S., Li, H., Fajardo, J.E., Fiser, A. and Birge, R.B. (2005) Transactivation of Abl by the Crk II adapter protein requires a PNAY sequence in the Crk C-terminal SH3 domain. *Oncogene*. **24**, 8187-99
30. Fiser, A., Filipe, S.R. and Tomasz, A. (2003) Cell wall branches, penicillin resistance and the secrets of the MurM protein. *Trends Microbiol.*, **11**, 547.
31. Hargitai, J., Lewis, H., Boros, I., Racz, T., Fiser, A., Kurucz, I., Benjamin, I., Vigh, L., Penzes, Z., Csermely, P. *et al.* (2003) Bimoclomol, a heat shock protein co-inducer, acts by the prolonged activation of heat shock factor-1. *Biochem.Biophys.Res.Comm.*, **307**, 689.
32. Lee, S.A., Shen, E.L., Fiser, A., Sali, A. and Guo, S. (2003) The zebrafish forkhead transcription factor Foxi1 specifies epibranchial placode-derived sensory neurons. *Development*, **130**, 2669.
33. Vernal, J., Fiser, A., Sali, A., Muller, M., Jose, C.J. and Nowicki, C. (2002) Probing the specificity of a trypanosomal aromatic alpha-hydroxy acid dehydrogenase by site-directed mutagenesis. *Biochem.Biophys.Res.Comm.*, **293**, 633.
34. Dosztanyi, Z., Magyar, C., Tusnady, G.E., Cserzo, M., Fiser, A. and Simon, I. (2003) Servers for sequence-structure relationship analysis and prediction. *Nucleic Acids Res.*, **31**, 3359.
35. Fiser, A. and Simon, I. (2000) Predicting the oxidation state of cysteines by multiple sequence alignment. *Bioinformatics.*, **16**, 251.
36. Fiser, A. and Simon, I. (2002) Predicting redox state of cysteines in proteins. *Methods Enzymol.*, **353**, 10.
37. Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B. and Sali, A. (2002) Reliability of assessment of protein structure prediction methods. *Structure.(Camb.)*, **10**, 435.
38. Simon, I., Fiser, A. and Tusnady, G.E. (2001) Predicting protein conformation by statistical methods. *Biochim.Biophys.Acta*, **1549**, 123.
39. Tudos, E., Fiser, A., Simon, A., Dosztanyi, Z., Fuxreiter, M., Magyar, C. and Simon, I. (2004) Noncovalent cross-links in context with other structural and functional elements of proteins. *J.Chem.Inf.Comput.Sci.*, **44**, 347.
40. Vetting, M.W., Hegde, S.S., Fajardo, J.E., Fiser, A., Roderick, S.L., Takiff, H.E. and Blanchard, J.S. (2006) Pentapeptide repeat proteins. *Biochemistry*, **45**, 1-10.
41. Zhan, C., Fedorov, E.V., Shi, W., Ramagopal, U.A., Thirumuruhan, R., Manjasetty,

- B.A., Almo, S.C., Fiser, A., Chance, M.R. and Fedorov, A.A. (2005) The ybeY protein from *Escherichia coli* is a metalloprotein. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **61**, 959-963.
42. Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Fiser A, Pazos F, Valencia A, Sali A, Rost B (2001) EVA: continuous automatic evaluation of protein structure prediction servers. *Bioinformatics* (2001) **17**(12) : 1242-3
 43. Chattopadhyay K, Bhatia S, Fiser A, Almo SC, Nathenson SG. (2006) Structural basis of inducible costimulator ligand costimulatory function: determination of the cell surface oligomeric state and functional mapping of the receptor binding site of the protein. *J Immunol* 177(6) : 3920-9
 44. Fiser A. (2004) Protein structure modeling in the proteomics era. *Expert Rev Proteomics*. **1**(1) : 97-110
 45. Rubinstein R, Fiser A
Predicting disulfide bond connectivity in proteins by correlated mutations analysis. *Bioinformatics* (2008) **24**(4) : 498-504
 46. Graslund S, Nordlund P, Weigelt J, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo, Fiser a. et al (2008) Protein production and purification. *Nature Methods* **5**(2) : 135-46